Solutions

Exercise 1

$$\begin{aligned}
\mathcal{L} &= \sum_{i=1}^{m} p_i L_i + \lambda(\sum_{i=1}^{m} 2^{-L_i} - 1) \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= \sum_{i=1}^{m} 2^{-Li} - 1 = 0 \Rightarrow \lambda = \frac{1}{ln2} \\
\frac{\partial \mathcal{L}}{\partial L_i} &= p_i + \lambda(-2^{-Li} ln2) = 0 \Rightarrow L_i = -log_2 p_i \\
&\Rightarrow min(E(L_x)) = \sum_{i=1}^{m} -p_i log_2 p_i = H(X)
\end{aligned}$$

But I don't know how to prove it is a minimum. It remains to be worked out.

Exercise 2

For a Golomb encoding, #bits taken for encoding each gap is

$$\begin{aligned}
L_i &= \left\lfloor \frac{ip}{ln2} \right\rfloor + 1 + \left\lceil log_2 \frac{ln2}{p} \right\rceil \\
&\le \frac{ip}{ln2} + 1 + log_2(\frac{ln2}{p} + 1) \\
&\le \frac{ip}{ln2} + 1 + log_2 \frac{3ln2}{p} \\
&= \frac{ip}{ln2} + 1 + log_2 \frac{1}{p} + log_2 3ln2
\end{aligned}$$

We want to prove that

$$\begin{aligned}
L_i &\le log_2 \frac{1}{p_i} + O(1) \\
&= log_2 \frac{1}{(1-p)^{i-1} p} + O(1) \\
&= log_2 (\frac{1}{1-p})^{i-1} + log_2(\frac{1}{p}) + O(1)
\end{aligned}$$

Compare both sides, we get

$$\begin{aligned}
\frac{ip}{ln2} + 1 + log_2 \frac{1}{p} + log_2 3ln2 &\le log_2(\frac{1}{1-p})^{i-1} + log_2(\frac{1}{p}) + O(1) \\
\frac{ip}{ln2} + 1 + log_2 3ln2 &\le log_2(\frac{1}{1-p})^{i-1} + O(1) \\
\frac{(i-1)p}{ln2} + \frac{p}{ln2} + 1 + log_2 3ln2 &\le (i-1) log_2 \frac{1}{1-p} + O(1)
\end{aligned}$$

Since $\frac{p}{ln2} + 1 + log_2 3ln2 < 6$, we only need to show

$$\begin{aligned}\frac{p}{ln2} &\leq log_2 \frac{1}{1-p} \\ p &\leq ln \frac{1}{1-p} \\ e^p &\leq \frac{1}{1-p} \\ e^{-p} &\geq 1-p\end{aligned}$$

which is given by the hint. Q.E.D

Exercise 3

We assume that $|L_j|$ is propotional to $\frac{1}{j}$, say $|L_j| = \frac{k}{j}$,

$$\begin{aligned}\sum_{j=1}^{m} |L_i| &= N \\ k \sum_{j=1}^{m} \frac{1}{j} &= N \\ \Rightarrow k &= \frac{N}{lnm + O(1)}\end{aligned}$$

The expected total number of bits required to gap-encode all the inverted lists is

$$\begin{aligned}E &= \sum_{j=1}^{m} (log_2 j + O(1))|L_i| \\ &= \sum_{j=1}^{m} |Li| log_2 j + O(N)\end{aligned}$$

We want to prove that

$$\sum_{j=1}^{m} |Li| log_2 j + O(N) \leq N \frac{log_2 m}{2} + O(N)$$

Since $|L_j| = \frac{k}{j}$, the above inequality becomes

$$\begin{aligned}\frac{N}{ln2(lnm + O(1))} \sum_{i=1}^{m} \frac{lnj}{j} + O(N) &\leq N \frac{log_2 m}{2} + O(N) \\ \frac{N}{ln2(lnm + O(1))} (\frac{ln^2 m}{2} + O(1)) + O(N) &\leq N \frac{log_2 m}{2} + O(N)\end{aligned}$$

Start from the left

$$
\begin{aligned}
\frac{N}{ln2(lnm+O(1))}(\frac{ln^2m}{2}+O(1))+O(N) &\leq \frac{N}{(ln2)(lnm)}(\frac{ln^2m}{2}+O(1))+O(N) \\
&\leq N\frac{log_2m}{2}+\frac{N}{(ln2)(lnm)}O(1)+O(N) \\
&\leq N\frac{log_2m}{2}+O(N)
\end{aligned}
$$

Q.E.D